

METHOD FOR ALIGNING GESTURE FEATURES OF IMAGE

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to the technical field of image recognition
5 and, more particularly, to a method for aligning gesture features of image.

2. Description of Related Art

In the field of gesture recognition, the vision-based static gesture
recognition is made possible by recognizing posture or shape of a gesture
image. Hence, techniques about extracting and matching gesture features
10 (e.g., posture or shape) of image are critical with respect to image
recognition.

Conventionally, curvature scale space (CSS) descriptors are utilized for
obtaining a quantitative shape feature description of the object for
recognizing gesture features of image, thereby providing a reliable feature
15 description even though the shape of the object being adversely affected by
size, rotation angle, and movement.

The conventional method for gesture recognition first captures an input
gesture image, and then a closed curve formed by a binary contour image of
the gesture image is determined by preprocessing the gesture image. A CSS
20 image of the gesture image is drawn based on the closed curve. Next, a
coordinate with a maximal peak in a coordinate-peak set formed by the CSS
image is selected as a basis point for alignment. A circular rotation is
performed to generate an aligned CSS image according to the basis point
for determining feature parameters of the gesture image. Finally, each

feature parameter of the plurality of sets of the gesture image is compared with each feature parameter of a plurality of reference gesture shapes represented as a basis point of the maximal peak, thereby determining a gesture shape corresponding to the gesture image.

5 However, the first several peaks have about equal values while exhibiting significant curvature changes in the CSS image of gesture shape due to rugged shape of static gesture in which peak may occur at each of three recessed portions between fingers except the thumb. Taking the gesture image with five fingers of a hand representing digit “5” as an
10 example, the maximal peak of the CSS image of the gesture image may occur at the recessed portion between the thumb and index finger, or at the recessed portion between the index and middle fingers. The CSS images represent the same gesture shape no matter where the maximal peak occurs, but the image recognizer may determine different results with the influence
15 of different maximal peaks. Furthermore, the image recognizer may also determine incorrect result when the CSS image is interfered by noise resulted in some other greater peaks. Due to the local limitation, the curvature can only record the “local” curved degree without labeling the size of the whole recessing or protruding area according to the whole
20 contour. Thus, the conventional image recognizer is unreliable and cannot directly determine the position of fingers in the image.

SUMMARY OF THE INVENTION

An object of the present invention is to provide a method for aligning

gesture features of image in which the basis point for alignment is selected based on the two-dimensional distribution of the coordinate-peak set instead of being selected from the coordinate with the maximal peak, thereby solving the location limitation and providing a reliable description
5 of gesture features.

Another object of the present invention is to provide a method for aligning gesture features of image in which curvature scale space is utilized to describe a gesture contour of image for preventing image shape from being adversely affected by size, rotation angle, and movement of image,
10 thereby providing a reliable description of gesture features.

To achieve the objects, the method for aligning gesture features of image of the present invention comprises the steps of: capturing an input gesture image; determining a closed curve formed by a binary contour image of the gesture image by preprocessing the gesture image; drawing a
15 curvature scale space (CSS) image of the gesture image based on the closed curve; performing a convolution operation with respect to the sequence of a coordinate-peak set formed by the CSS image and a predefined function $F(x)$ to designate the coordinate with maximal value of integration as a basis point for obtaining feature parameters of the gesture image; and comparing
20 each feature parameter of the gesture image with each feature parameter of a plurality of reference gesture shapes for determining a gesture shape corresponding to the gesture image.

Other objects, advantages, and novel features of the invention will become more apparent from the following detailed description when taken

in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a flow chart illustrating a process of aligning gesture features of image according to the invention;

5 FIG. 2 schematically illustrating a circle of curvature according to the invention;

FIG. 3 is a flow chart illustrating a process of calculating curvature scale space image according to the invention;

10 FIG. 4 is a schematic drawing of binary contour image of the gesture image according to the invention;

FIG. 5 is a chart depicting the curvature scale space image of FIG. 4;

FIG. 6 is a flow chart illustrating a process of calculating feature parameters of the gesture image according to the invention;

15 FIG. 7 is a schematic drawing of the predefined function according to the invention;

FIG. 8 is a schematic drawing of the aligned curvature scale space image according to the invention; and

20 FIG. 9 is a flow chart illustrating a process of comparing feature parameter of input gesture image with feature parameter of predetermined reference gesture shape according to the invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

With reference to FIG. 1, there is shown a process of aligning gesture features of image in accordance with preferred embodiment of the present invention. At first, an input gesture image is captured (step S101). Because

a gesture feature is recognized based on a contour of gesture image in the present invention, the next step is to obtain a closed curve formed by binary contour image of the input gesture image by dividing the input gesture image by means of image preprocessing technique (step S102). The
5 technique of calculating binary contour image is well known, and thus a detailed description is deemed unnecessary.

The process further enables an image processing device to draw a curvature scale space (CSS) image of gesture contour image based on the closed curve (step S103). Then, a convolution operation is performed with
10 respect to the sequence of a coordinate-peak set formed by the CSS image and a predefined function $F(x)$ in order to obtain a basis point for alignment according to the two-dimensional distribution of the coordinate-peak set (step S104), where $F(x) \geq 0$ and $F(x)$ is an asymmetric function. Next, a coordinate with maximal value of integration is designated as the basis
15 point for obtaining feature parameters of the gesture image (step S105).

For the purpose of recognizing a gesture feature represented by feature parameter F^I , there is provided a database including a plurality of feature parameters F^S corresponding to reference gesture shapes for being respectively compared with the feature parameter F^I . In general, a static
20 gesture can be used to represent a certain digit. For example, the index finger represents digit "1", the index finger plus middle finger represent digit "2", the five fingers of a hand represent digit "5", and so on. In this embodiment, the database contains various shapes of reference gesture, including different sizes, movements, and rotation angles of gesture, for

representing different digits. In addition, the database may contain other various shapes of static gesture such as sign language gesture.

The process further comprises a final step of comparing each parameter F^I of the gesture image with each parameter F^S of reference gesture shapes by utilizing a nearest neighbor algorithm to recognize gesture for finding a nearest reference gesture shape with respect to input gesture image so as to determine a gesture shape corresponding to the input gesture image (step S106).

With reference to FIG. 3 and FIG. 2, the detailed process of the step S103 for calculating CSS image is illustrated. In FIG. 2, curvature κ at point P of closed curve Γ is defined as either gradient angle of tangent ϕ with respect to arc length parameter u on the instant of time or an inverse of radius r of circle of curvature R . Thus, curvature κ can be expressed in the following equation:

$$\kappa = \frac{d\phi}{du} = \frac{1}{r}.$$

Also, closed curve Γ can be expressed as $\{x(u), y(u)\}$ in terms of arc length parameter u in which u is normalized to have a value between 0 and 1. After rearrangement, above curvature κ of closed curve Γ can be expressed as follows:

$$\kappa(u) = \frac{\dot{x}(u)\ddot{y}(u) - \ddot{x}(u)\dot{y}(u)}{(\dot{x}^2(u) + \dot{y}^2(u))^{\frac{3}{2}}} \text{ (step S301),}$$

$$\text{where } \dot{x}(u) = \frac{dx}{du}, \ddot{x}(u) = \frac{d^2x}{du^2}, \dot{y}(u) = \frac{dy}{du} \text{ and } \ddot{y}(u) = \frac{d^2y}{du^2}.$$

Next, a convolution operation is performed with respect to closed curve Γ and an one-dimensional Gauss function $g(u, \sigma)$ for obtaining a smooth curvature function $\Gamma_\sigma = \{X(u, \sigma), Y(u, \sigma)\}$ and its curvature:

$$\kappa(u, \sigma) = \frac{X_u(u, \sigma)Y_{uu}(u, \sigma) - X_{uu}(u, \sigma)Y_u(u, \sigma)}{(X_u(u, \sigma)^2 + Y_u(u, \sigma)^2)^{3/2}} \quad (\text{step S302}),$$

5 where σ is standard deviation. The Gauss function can be expressed as follows:

$$g(u, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{u^2}{2\sigma^2}\right),$$

and $X(u, \sigma)$ and $Y(u, \sigma)$ in the smooth curvature function Γ_σ can be expressed respectively as follows:

$$10 \quad X(u, \sigma) = x(u) * g(u, \sigma) = \int_{-\infty}^{\infty} x(v) \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{(u-v)^2}{2\sigma^2}\right) dv,$$

$$Y(u, \sigma) = y(u) * g(u, \sigma) = \int_{-\infty}^{\infty} y(v) \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{(u-v)^2}{2\sigma^2}\right) dv.$$

Moreover, $X_u(u, \sigma)$, $X_{uu}(u, \sigma)$, $Y_u(u, \sigma)$ and $Y_{uu}(u, \sigma)$ of curvature $\kappa(u, \sigma)$ are expressed respectively as follows:

$$X_u(u, \sigma) = x(u) * \dot{g}(u, \sigma), \quad X_{uu}(u, \sigma) = x(u) * \ddot{g}(u, \sigma),$$

$$15 \quad Y_u(u, \sigma) = y(u) * \dot{g}(u, \sigma), \quad Y_{uu}(u, \sigma) = y(u) * \ddot{g}(u, \sigma),$$

$$\text{where } \dot{g}(u, \sigma) = \frac{\partial}{\partial u} g(u, \sigma) \text{ and } \ddot{g}(u, \sigma) = \frac{\partial^2}{\partial u^2} g(u, \sigma).$$

Generally, a smooth curve has a higher resolution when σ is relatively small while having a lower resolution when σ is relatively large. Hence, the contour drawn by $\Gamma_\sigma = \{X(u, \sigma), Y(u, \sigma)\}$ is more smooth as σ

increases. Finally, different standard deviations σ are used to find a location having zero curvature in $\Gamma_\sigma = \{ X(u, \sigma), Y(u, \sigma) \}$ and thus all locations having zero curvature are drawn under different standard deviations σ (S303). As a result, a CSS image of input gesture image
5 contour is obtained.

With reference to FIG. 4, there is shown a schematic drawing of a binary contour image of the input gesture image I. Based on the above process, the $u - \sigma$ coordinate I1 depicting the input gesture image I shown in FIG. 4 is determined as illustrated in FIG. 5, where axis of abscissa is u
10 which is normalized to have a value between 0 and 1 and axis of ordinate is σ . In the $u - \sigma$ coordinate, the position of $\kappa(u, \sigma) = 0$ is defined as zero curvature point.

In response to the drawing of CSS image of input gesture I, an extraction of feature parameters F^I of input gesture image I is calculated
15 based on a process illustrated in the flow chart of FIG. 6. First, all sets of peaks in CSS image are calculated to form a set of coordinate-peak (step S601), which is denoted as $u - \sigma$ set, and expressed as

$$\{(u_i, \sigma_i)\}_{i=1, \dots, N}^{\text{Original}} = (u_1, \sigma_1), (u_2, \sigma_2), \dots, (u_j, \sigma_j), \dots, (u_N, \sigma_N),$$

where N is the number of all detected peaks in CSS image. Next, a basis
20 point k_o is selected based on the two-dimensional distribution of the coordinate-peak set (step S602) expressed as:

$$k_o = \arg \max_k \left(\sum_{i=1}^{k-1} \sigma_i \cdot F(1 + u_i - u_k) + \sum_{i=k}^N \sigma_i \cdot F(u_i - u_k) \right),$$

so as to solve the local limitation of the prior method for selecting

coordinate with the maximal peak.

In the aforementioned function for selecting the basis point, $F(x)$ is a zero-mean Gauss function having a value between 0 and 1, for example,

$$F(x) = \frac{e^{-[x^2/2\sigma^2]}}{\sqrt{2\pi}\sigma}, \quad 0 \leq x < 1, \text{ where } \sigma \text{ is defined for controlling the changing}$$

5 rate of $F(x)$ in the period of $[0,1]$. In order to reduce the probability of obtaining multiple identical maximal values, the value field of $F(x)$ is defined as $F(x) \geq 0$, where $F(x)$ is an absolute increasing or decreasing asymmetric function, or a non-continuous function. In this embodiment, $F(x)$ is an absolute decreasing function as illustrated in FIG. 7. Since
10 zero-mean Gauss function is a one-to-one function in the period of $[0,1]$, the probability of obtaining multiple basis points k_0 of functions, such as surge function, step function, or the like, if the inputted coordinate-peak set is non-periodical or symmetric. The function used for selecting basis points can be taken as performing a convolution operation with respect to
15 $\{(u_i, \sigma_i)\}_{i=1, \dots, N}^{\text{Original}}$ and $F(x)$ so as to extract the coordinate u_{k_0} with the maximal value of integration from the $u - \sigma$ coordinate I1.

Then, the $u - \sigma$ coordinate I1 can be expressed as the following coordinate-peak set after being aligned through a circular rotation with respect to u_{k_0} (i.e. basis point) to obtain an aligned $u - \sigma$ coordinate I2 as
20 shown in FIG. 8 (step S603):

$$\{(u_i, \sigma_i)\}_{i=1, \dots, N}^{\text{aligned}} = \{(0, \sigma_{k_0}), (u_{k_0+1} - u_{k_0}, \sigma_{k_0+1}), \dots, (1 + u_1 - u_{k_0}, \sigma_1), \dots, (1 + u_{k_0-1} - u_{k_0}, \sigma_{k_0-1})\},$$

in which $u = 0$ is the position of the maximal peak corresponding to u -axis.

As a result, feature parameter $F^I = \{(u_i^I, \sigma_i^I)\}_{i=1, \dots, N}^{\text{aligned}}$ of the input gesture image I can be calculated according to the aligned $u - \sigma$ coordinate I2 (step S604), where N is the number of all of the detected peaks in the CSS image, I is the input gesture image I.

5 After obtaining the feature parameter F^I of the input gesture image I, the step S106 as shown in FIG. 1 is performed to compare F^I with feature parameter F^S of reference gesture shape. As a result, a possible similarity between F^I and F^S is obtained to determine a corresponding gesture shape of input gesture image I. The feature parameter F^S of predetermined
10 reference gesture shape in the database can be expressed as $F^S = \{(u_j^S, \sigma_j^S)\}_{j=1, \dots, M}^{\text{aligned}}$, where M is the number of peaks of reference gesture shape, and S is the reference gesture shape.

With reference to FIG. 9, there is shown a flow chart illustrating a process of comparing F^I with F^S . First, a distance function is employed to
15 obtain a sum of distance of the matched peaks and distance of the unmatched peaks between F^I and F^S (step S901). The distance function of this embodiment can be expressed as follows:

$$\text{dist}(F^I, F^S) = \sum_{\text{matched peaks}} \sqrt{(u_i^I - u_j^S)^2 + (\sigma_i^I - \sigma_j^S)^2} + \sum_{\text{unmatched peaks}} \sigma_i^I + \sum_{\text{unmatched peaks}} \sigma_j^S,$$

which indicates that a small distance refers to a high similarity of two
20 shapes.

Finally, a nearest neighbor algorithm is utilized to determine a reference gesture shape of the input gesture image I for gesture recognition (step S902). Therefore, the image processing device can follow the

aforesaid functions to find out the nearest reference gesture shape with respect to input gesture image I in the process of recognition

In view of the foregoing, the method for aligning gesture features of image according to the present invention can directly determine the position
5 of fingers in the image. This invention can also find out the areas having greater recessed portions by utilizing the tendency of combining the process of gradually smoothing the image and the CSS data corresponding to the image. As for the gesture image, the area having greater recessed portion is one of the possible recessed portions between fingers, thus those more
10 obvious recessed portions in the gesture image can be figured out according to the present invention.

It is known that the invention ensures the extracted CSS image to be reliable after being aligned. The invention can be applied in games, human computer interfaces, sign language recognition, video surveillance, image
15 and video retrieval, etc. Particularly, the invention is suitable in game related applications which require features of gesture recognition, smooth manipulation, highly interaction between players and game, to greatly increase entertainment effect.

Although the present invention has been explained in relation to its
20 preferred embodiment, it is to be understood that many other possible modifications and variations can be made without departing from the spirit and scope of the invention as hereinafter claimed.